

FUSÃO DE CARACTERÍSTICAS ADAPTATIVA POR CONSULTA PARA DIVERSIFICAÇÃO VISUAL

Felipe Souza Cordeiro¹; Fabiana Cristina Bertoni²; Rodrigo Tripodi Calumby³

1. Graduando em Engenharia de Computação, Universidade Estadual de Feira de Santana, fscordeiro@ecomp.uefs.com.br

2. Orientador, Departamento de Ciências Exatas, Universidade Estadual de Feira de Santana, fbertoni@gmail.com

3. Coorientador, Departamento de Ciências Exatas, Universidade Estadual de Feira de Santana, rtcalumby@ecomp.uefs.br

PALAVRAS-CHAVE: Aprendizado de Máquina, Diversidade Visual, Recuperação de Imagens

INTRODUÇÃO

Com o aumento da popularização das tecnologias fica cada vez mais fácil adquirir um dispositivo capaz de capturar e armazenar dados. Esses dados são utilizados das mais diversas formas, entre eles, a utilização em redes sociais e sites, gerando a cada dia uma grande quantidade de conteúdos digitais multimídia. Estes conteúdos são explorados para várias áreas, como: sensoriamento remoto e análise forense. Logo, é preciso encontrar maneiras eficientes de retornar resultados relevantes às necessidades de cada usuário.

Na recuperação de imagens, por exemplo, pode-se utilizar de informações textuais (metadados, palavras-chaves, etc) e descrições das propriedades visuais (cor, forma, textura, etc). No processo de recuperação de imagens, a relevância depende das intenções de busca do usuário. É comum que os descritores de conteúdo individualmente não representem da melhor maneira o conteúdo semântico de uma imagem. Esse problema é conhecido como gap semântico. Visando amenizar este problema, descritores são combinados para adaptar as buscas às necessidades dos usuários (Atrey *et al* (2010), Calumby *et al* (2014)). É preciso definir dinamicamente quais as melhores características que representam melhor uma consulta ou o nível de importância que cada característica disponível pode ter na definição de relevância de cada item da coleção de dados.

Há dois tipos de técnicas para fusão de características: *early fusion* e *late fusion*. A abordagem *early fusion* primeiro extrai as características referentes a cada uma das propriedades para então as combinar. As abordagens baseadas em *late fusion* realizam a extração das características para cada uma das propriedades, mas os algoritmos de aprendizado são aplicados a cada uma das propriedades em separado. O resultado é a classificação dos itens da base.

O objetivo deste trabalho foi desenvolver um modelo de diversificação visual baseado em fusão de características de imagens. Foi utilizado o método de Algoritmos Genéticos (AG) (Goldberg, 1989) para desenvolvimento dinâmico e adaptativo por consulta, do modelo proposto, a fim de apresentar uma solução capaz de definir dinamicamente a importância que cada característica possui na combinação dos descritores para recuperação multimodal, visual e textual.

METODOLOGIA

Para definir dinamicamente quais são as melhores características que representam uma consulta ou a importância que cada característica possui, o projeto propôs a utilização da técnica de AG como aprendizado de máquina de forma a otimizar um modelo de diversificação visual baseado em fusão de características. Para tal, foi preciso modelar e implementar um método de diversificação no framework Diversify (Cordeiro, 2014)

e avaliar seus resultados comparando-os com outras abordagens de diversificação como métodos de agrupamento, Calumby *et al* [(2014), (2015)]. A comparação entra as abordagens já existentes e a proposta por nós foi realizada na tarefa de busca multimodal de imagens e sumarização visual, com base na avaliação proposta pelo MediaEval Diverse Social Images Task, (Ionescu *et al*, (2015)).

O objetivo dos AG é resolver problemas específicos através de evoluções, como a evolução natural da biologia. Assim, esses algoritmos simulam processos naturais de sobrevivência e reprodução das populações, essenciais em sua evolução. Como na seleção natural, os indivíduos mais aptos (melhores soluções) terão chances de sobreviver e gerar descendentes, ao contrário dos indivíduos menos aptos. O AG utiliza de alguns elementos, como: população de indivíduos, compostos por genes, sendo um gene um parâmetro codificado do indivíduo. Os genes podem sofrer mutações e cruzamentos, de maneira probabilística, além de poder utilizar o operador de seleção e replicação de indivíduos de maneira direta de uma geração para outra. E o fitness, o valor de fitness é a aptidão do indivíduo no meio.

O método de fusão proposto utiliza o AG, para encontrar os pesos de cada característica utilizada no cálculo da distância entre as imagens. Para o cálculo da distância é utilizada a Equação 1, sendo A, B, C e D os pesos encontrados e D1, D2, D3 e D4 as distâncias entre as imagens i e j utilizando cada descritor. A partir das distâncias entre as imagens são gerados os agrupamentos. Então, os grupos são avaliados com métricas de índices de qualidade. A medida utilizada no processo de avaliação dos grupos foi a Silhueta (Rousseeuw, 1987), essa avaliação é o fitness do processo do algoritmo genético.

$$Dist(i,j) = \frac{A(D1_{ij}) + B(D2_{ij}) + C(D3_{ij}) + D(D4_{ij})}{A+B+C+D} \quad (1)$$

A Silhueta fornece uma medida de como cada objeto se encontra dentro de seu grupo, para cada objeto o valor de Silhueta é calculado e esse valor varia de -1 a 1. Quanto mais próximo a 1, o objeto provavelmente está no grupo correto. Um valor próximo a 0, é indefinido, isto é, não se pode deduzir se ele está no grupo correto ou não e próximo a -1, o objeto provavelmente está no agrupamento errado.

O AG implementado neste trabalho utilizou o framework JMetal (Durillo & Alba, (2010)). E o método foi experimentado na base de dados fornecida pela coleção da tarefa do *MediaEval Diverse Social Images Task*. O conjunto de desenvolvimento (devset) contém 70 consultas (20,757 fotos do Flickr). O conjunto de test (testset) apresenta 65 consultas (19,017 fotos do Flickr). A avaliação provida pela tarefa é com base no desempenho, tanto para a diversidade como para a relevância. As seguintes métricas são computadas: *Cluster Recall* em X (CR@X) - uma medida que avalia quantos grupos diferentes do GT estão representados entre os X resultados encontrados no topo da lista final de recuperação das imagens (somente imagens relevantes são consideradas); *Precision* em X (P@X) - mede o número de fotos relevantes entre os X resultados encontrados no topo da lista final de recuperação das imagens; e *F1-Measure* em X (F1@X) - é a média harmônica dos dois anteriores.

Essa base também fornece um gabarito, *Ground-Truth* (GT), que serve de parâmetro para validar a qualidade dos agrupamentos gerados. Para análise da qualidade da medida de Silhueta no nosso problema foram gerados grupos perfeitos com base no gabarito fornecido pela coleção da tarefa citada anteriormente. Os agrupamentos foram então avaliados com base na medida de Silhueta utilizando cada descritor disponível individualmente.

Após avaliar os descritores isoladamente utilizando a Silhueta como medida de *fitness* e a pureza para quantificar quantas imagens um cluster x tem dos objetos de uma partição do cluster y

do GT, foram selecionados os 10 melhores descritores para as modalidades visual e multimodal; e os 9 descritores textuais disponíveis.

RESULTADOS

Foi implementado um módulo do AG que permite utilizar as funcionalidade já disponibilizadas pelo JMetal ou implementar novas técnicas na utilização do AG. Essas técnicas tem seus comportamentos influenciados pelos parâmetros, assim, uma pequena mudança de parâmetro já é o suficiente para modificar o processo de descoberta das soluções, o que afeta o processo de agrupamento, sendo possível melhorar ou não os resultados. Visando encontrar os parâmetros ideais, foram avaliados algumas possibilidades de configuração, variando um parâmetro e fixando os outros.

Após as execuções variando os parâmetros, citados na seção da metodologia, tais como, taxa de mutação e cruzamento e número de indivíduos da população, não foram encontrados padrões nos resultados, as configurações em alguns casos deram resultados muito próximos. A Tabelas 1 ilustra os melhores resultados encontrados para cada modalidade, visual, textual e multimodal, respectivamente, apresentando uma comparação com os resultados gerados de forma não diversificada, utilizando o K-Medóides com fusão de características pela médias dos descritores e a nossa abordagem.

Tabela 1. Tabela comparativa com diferentes métodos de agrupamentos.

		Visual		Textual		Multimodal	
Métricas	NoDiv	K-Medóides	AG	K-Medóides	AG	K-Medóides	AG
P@20	0,5555	0,5102	0,5312	0,5438	0,5461	0,5336	0,5328
CR@20	0,3609	0,3885	0,3965	0,4006	0,3970	0,4055	0,3925
F1@20	0,4131	0,4213	0,4320	0,4423	0,4390	0,4394	0,4339

Os resultados foram obtidos utilizando os melhores descritores encontrados para cada modalidade, sendo os 10 melhores para a modalidade visual: Col, Lum, Ophist, Gabor, Lbp, Tamura, Rilbp, Gist, Jhist e Sch. Para a modalidade textual foram utilizados: BM25, BM25Orig, BM25Ext, COSINE, COSINE_ME, TFIDF, JACCARD, TFIDF_MEProd e TFIDF_ME. E os 10 melhores descritores para modalidade multimodal, visual e textual, foram: Col, Lum, Ophist, Gabor, Lbp, Tamura, Rilbp, COSINE, COSINE_ME e Sch. A Figura 1 apresenta os 18 melhores descritores encontrados de acordo com a medida de Silhueta.

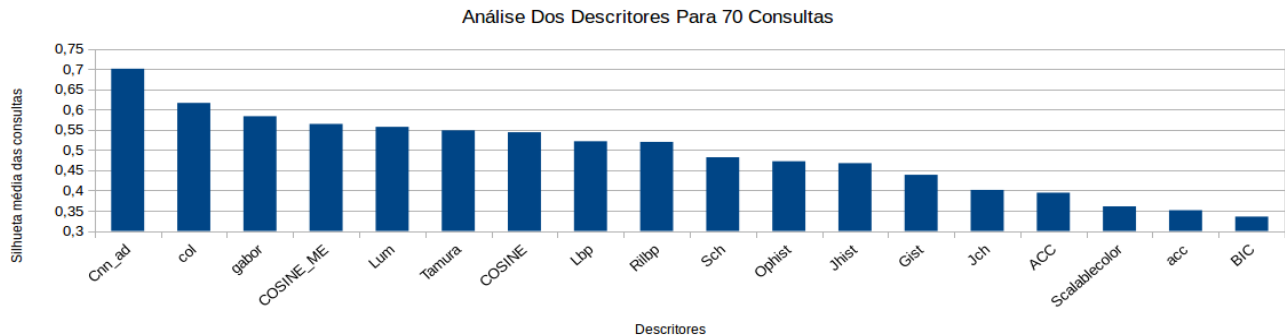


Figura 1: Descritores ordenados por ordem decrescente da Silhueta média utilizando 70 consultas.

CONCLUSÃO

Este trabalho descreveu o desenvolvimento e implementação de um método de fusão de características para o processo de diversificação de imagens. Também foram apresentados os conceitos utilizados no desenvolvimento do método proposto. O desenvolvimento modularizado permite obter liberdade para a experimentação das diferentes possibilidades de configurações paramétricas. Os melhores resultados obtidos com as configurações testadas são ilustradas na Tabela 1. Utilizando a abordagem proposta para as modalidades textual e multimodal não houve melhora em relação a utilização do K-Medóides, entretanto, observou-se uma pequena melhoria na abordagem puramente visual. É preciso ser observado que ainda existem diversas possibilidades de configuração a serem combinadas e testadas, esses novos testes podem apresentar uma melhora significativa nesta nova abordagem de fusão de características.

A abordagem proposta possibilita a recuperação de imagens de maneira dinâmica, de forma a descobrir para cada consulta quais descritores devem possuir maior relevância. Essa abordagem é inovadora, pois durante as pesquisas não foram encontradas pesquisas utilizando essa perspectiva de pesos dinâmicos a serem evoluídos de forma a encontrar a melhor fusão entre eles.

REFERÊNCIAS

- B. Ionescu, A. L. Gînscă, B. Boteanu, A. Popescu, M. Lupu, H. Müller. Retrieving Diverse Social Images at MediaEval 2015: Challenge, Dataset and Evaluation.
- F. S. Cordeiro, V. P. Santana, and R. T. Calumby, "Diversify - um micro arcabouço para avaliação de métodos de sumarização visual," XVIII Seminário de Iniciação Científica da UEFS, 2014.
- G. Chiachia, M. A. Gonçalves, R. da S. Torres. Recod @ MediaEval 2015: Diverse Social Images Retrieval 2015 (Working Notes Proceedings of the MediaEval Workshop).
- Goldberg, D. E. Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, 1989.
- J.J. Durillo, A.J. Nebro, E. Alba. JMetal. Disponível em: <<http://jmetal.sourceforge.net/>>. Acessado em: 26 de Agosto de 2016.
- P. Atrey, M. Hossain, A. El Saddik, and M. Kankanhalli. Multimodal fusion for multimedia analysis: a survey. Multimedia Systems, 16(6), 1–35 (2010).
- Rousseeuw, Peter J. "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis." Journal of computational and applied mathematics 20 (1987): 53-65.
- R. T. Calumby, R. da S. Torres e M. A. Gonçalves. Multimodal Retrieval with Relevance Feedback Based on Genetic Programming. Multimedia Tools and Applications, 69(3):991-1019, 2014.
- R. T. Calumby, V. P. Santana, F. S. Codeiro, O. A. B. Penatti, L. T. Li, G. Chiachia, R. da S. Torres. Recod @ MediaEval 2014: Diverse Social Images Retrieval. Diverse Social Images Retrieval 2014. Working note papers.
- R. T. Calumby, I. B. A. C. Araujo, V. P. Santana, J. A. V. Munoz, O. A. B. Penatti, L. T. Li, J. Almeida; G. Chiachia, M. A. Gonçalves, R. da S. Torres. Recod @ MediaEval 2015: Diverse Social Images Retrieval 2015. Working Notes papers.